


# Chapter 9

## Comparing Big Data Analysis Techniques

Santosh Ramkrishna Durugkar

 <https://orcid.org/0000-0002-5079-2224>

Independent Researcher, India

### ABSTRACT

*Big data refers to the large volume of data which can be processed to get the relevant, required, and meaningful data within less time. It is required to apply sophisticated methods and techniques to process big data. There are many methods and techniques available, like regression analysis, time series analysis, sentiment analysis, descriptive analysis, predictive analysis, association analysis with sampling, machine learning, visualization techniques, classification, and qualitative and quantitative analysis. In the future, there is a need to enhance the performance of these techniques due to increasing size of data. Many applications are based on any of these techniques to process the large volume of data in order to retrieve the meaningful data. It is expected, big data analysis techniques will filter and process the large volume and find the relevant one. Analysis of big data is very helpful in many areas like businesses, industries, and other sectors.*

### 1. INTRODUCTION

Analyzing the big data is nothing but processing *raw data* to get meaningful and relevant data. Retrieved meaningful data can be useful in many applications which are based on different approaches like decision making, probabilistic analysis, and deterministic analysis. With the help of existing data analysis techniques, one can form the business strategies like *identifying customer segments, customer retention, possible future business growth etc.* However, it is suggested one can use data analysis technique as per the requirement i.e. type of input data and insights to uncover.

With the increasing size of users and applications data size is also increasing & in future it could be more than expected. Therefore, in future there would be a requirement of machines with good configurations & best analysis techniques. There are many existing analysis techniques available like *regression analysis, time series analysis, sentiment analysis, descriptive analysis, predictive analysis, diagnostic*

DOI: 10.4018/979-8-3693-0413-6.ch009

*analysis, association analysis with sampling, ensemble method for analysis, machine learning, visualization techniques, classification, and genetic algorithms and qualitative & quantitative analysis.*

Big data refers to the large volume of data and can be characterized on the basis of *volume, variety and velocity* (V3 model of big data). Volume refers to the quantity of data and keeping the storage as per the requirement. Variety refers to the different data i.e. structured and unstructured data and can have different forms. Velocity refers to the keeping an eye on incoming data with the ever increasing size of users and applications. Big data analysis techniques must deal with the metadata i.e. data about data. Metadata reveals the quality, and characteristics about the data. To process and retrieve the meaningful data, it is expected that analysis techniques must be capable enough to process the metadata too. Identifying the ‘*relevance*’ among the different segments (different variables) data is very important & known as ‘*association rule*’ mining. It is also expected to process the real time data (RT data), qualitative and quantitative data in order to get the relevant data from large volume. Big data refers to processing the ‘n’ large datasets & more sophisticated big data analysis techniques can be applied to extract the relevant data with less processing time. Examples of big data are as follows – *server web logs, patient records in hospitals, data related to surveillance, institutional data, large time series data, audio and video archives, and the most popular datasets from ‘e-commerce’ industries.*

### 1.1 Challenges in Big Data Analysis

Following are the challenges in big data analysis –

- *Storage*
- *Security*
- *Quality*
- *Scalability*
- *Analysis issues*
- *Selection of appropriate tool to analyze the large volume of data*
- *Lack of trained people*

## 2. LITERATURE REVIEW

In the research paper (Khurana 2014; Berisha et al. 2022) author has pointed the change in technological era i.e. use of cloud computing and big data and these two technologies are discussed in this paper. As the term define bi data, it deal with few steps such as, collecting the raw data, processing it, forming and handling the queries etc. These technologies are rapidly changing the face and pace of data driven applications. Cloud computing which refers to the infrastructure, hardware, software where the data will reside. There are public and private cloud service providers and this technology can be divided into the infrastructure as a service, platform as a service, and software as a service. Nowadays, AWS amazon web services are changing the face of data driven applications. Along with Amazon there are many cloud service providers such as Google, Microsoft etc. In this paper (Cheng et al. 2015; Attaallah et al, 2022) authors have focused on integrity and security of the data stores on the cloud. To protect this data one can divide it in multiple parts and store on multiple locations. Securing the big data on cloud authors have proposed a trapdoor function using various data elements to analyze, compare and simulate the

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/comparing-big-data-analysis-techniques/336352](http://www.igi-global.com/chapter/comparing-big-data-analysis-techniques/336352)

## Related Content

---

### A Business Intelligence Project-Oriented Course: A Breast Cancer Research Case

Dima Alberg (2019). *International Journal of Business Intelligence Research* (pp. 29-35).

[www.irma-international.org/article/a-business-intelligence-project-oriented-course/232238](http://www.irma-international.org/article/a-business-intelligence-project-oriented-course/232238)

### Weighting Imputation for Categorical Data

Liang-Ting Tsai, Chih-Chien Yang and Timothy Teo (2014). *Encyclopedia of Business Analytics and Optimization* (pp. 2706-2716).

[www.irma-international.org/chapter/weighting-imputation-for-categorical-data/107449](http://www.irma-international.org/chapter/weighting-imputation-for-categorical-data/107449)

### A Prescriptive Stock Market Investment Strategy for the Restaurant Industry using an Artificial Neural Network Methodology

Gary R. Weckman, Ronald W. Dravenstott, William A. Young II, Ehsan Ardjmand, David F. Millie and Andy P. Snow (2016). *International Journal of Business Analytics* (pp. 1-21).

[www.irma-international.org/article/a-prescriptive-stock-market-investment-strategy-for-the-restaurant-industry-using-an-artificial-neural-network-methodology/142778](http://www.irma-international.org/article/a-prescriptive-stock-market-investment-strategy-for-the-restaurant-industry-using-an-artificial-neural-network-methodology/142778)

### Chaos in Economics

Wei-Bin Zhang (2014). *Encyclopedia of Business Analytics and Optimization* (pp. 440-460).

[www.irma-international.org/chapter/chaos-in-economics/107248](http://www.irma-international.org/chapter/chaos-in-economics/107248)

### All About Analytics

Hugh J. Watson (2013). *International Journal of Business Intelligence Research* (pp. 13-28).

[www.irma-international.org/article/all-analytics/76909](http://www.irma-international.org/article/all-analytics/76909)